ARTICLE IN PRESS

Displays xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Sound-of-Tapping user interface technology with medium identification☆

Seok-Jeong Song, Hyoungsik Nam*

Department of Information Display, Kyung Hee University, Seoul 02447, Republic of Korea

ARTICLE INFO

Keywords:
Sound-of-tapping
User interface
Support vector machine
Medium identification

ABSTRACT

Touch screen technologies have become the mainstream of input interfaces for mobile devices. However, due to the small size screens, it is difficult to offer more than two inputs at the same time. In order to achieve input richness, this paper proposes a new sound-of-tapping (S-TAP) technology that is a novel user interface utilizing only microphones embedded in most mobile devices. S-TAP technology extracts the positions of tapping sounds generated at the outside of a smartphone by means of frequency domain features and support vector machine (SVM). The error ratio of an S-TAP algorithm implemented in a smartphone is measured as 8.6%. Additionally, this technology enables the smartphone to identify mediums that the user is tapping on. In the end, we demonstrate two smartphone applications to scroll web pages only by S-TAP and to zoom and rotate a map image by S-TAP as well as touch screen.

1. Introduction

Mobile devices such as tablet PCs, smartphones, and smart watches, have been significantly evolved with various cutting-edge technologies. Especially in smartphones, substantial advancement has been achieved every year. The screens feature the high resolution of 2560 × 1440 as well as multi touch support with active matrix liquid crystal display (LCD) or active matrix organic light emitting diode (AMOLED) display. The powerful central processing units (CPUs) of octa-core structure and 2.15 GHz clock speed are integrated along with 3 GB random access memory (RAM) and separate graphic processing unit (GPU). The cameras provide high resolution images and videos of 16 and 8 mega pixels, respectively. Wireless interconnections are established through Bluetooth, wireless LAN, and near field communication (NFC). In addition, a variety of sensors such as accelerometer, barometer, digital compass, fingerprint sensor, global positioning system (GPS), and gyroscope are embedded to diversify the applications of mobile devices.

To easily access and manipulate various functions in small size mobile devices, the portable input tools are preferred such as fingers and pencils rather than bulky keyboards and mouses. Until now, touch screen technologies have become the mainstream of input interfaces in mobile applications [1–3]. However, it is difficult for current touch interfaces to offer more than two inputs simultaneously due to small size screens [4,5]. Therefore, many researchers have been struggling to leverage the other input approaches for input richness. The one way is to allow multiple touch input tools to be separately sensed and distinguished on the screen at the same time [6–9]. Another method is to use other embedded sensors such as gyroscope, accelerometer, magnetometer, and image sensor to expand input modalities of mobile devices [10–14].

This paper describes a sound-of-tapping (S-TAP) technology that is a new user interface to use only microphones embedded in most mobile devices like a smartphone shown in Fig. 1. Human beings can find directions and positions of audio signals with one or two ears. In the same way, mobile devices with microphones can also extract direction and position information from received audio signals. Input sounds take place by tapping with fingers on tables or desks where the smartphone is placed. Then, from tapping sounds recorded by microphones, the position information is extracted with a proposed classification method that employs a support vector machine (SVM) algorithm [15,16]. SVM is one of popular machine-learning tools in the area of classification, especially, pattern recognition. It is a very efficient classifier in terms of classification accuracy, computational time, and stability [17] compared to other algorithms such as Gaussian mixture model, K-nearest neighbor, and radial basis function neural network. Furthermore, SVM can implement a non-linear classification using various kinds of kernel functions [15] that are polynomial, Gaussian, radial basis, and sigmoid functions for example. The recognized position is exploited to expand the input expressivity along with existing touch screen technologies.

The rest of the paper is organized as follows: Existing researches about several input modalities for mobile devices are reviewed in Section 2. The signal extraction method and classification scheme for a proposed S-TAP technique are described in Section 3. Section 4 presents evaluation results of the proposed S-TAP algorithm by means of

☆ This paper was recommended for publication by Pen-Cheng Wang.

* Corresponding author.
E-mail address: hyoungsiknam@khu.ac.kr (H. Nam).

https://doi.org/10.1016/j.displa.2018.02.003
Received 21 March 2017; Received in revised form 6 February 2018; Accepted 8 February 2018
0141-9382/

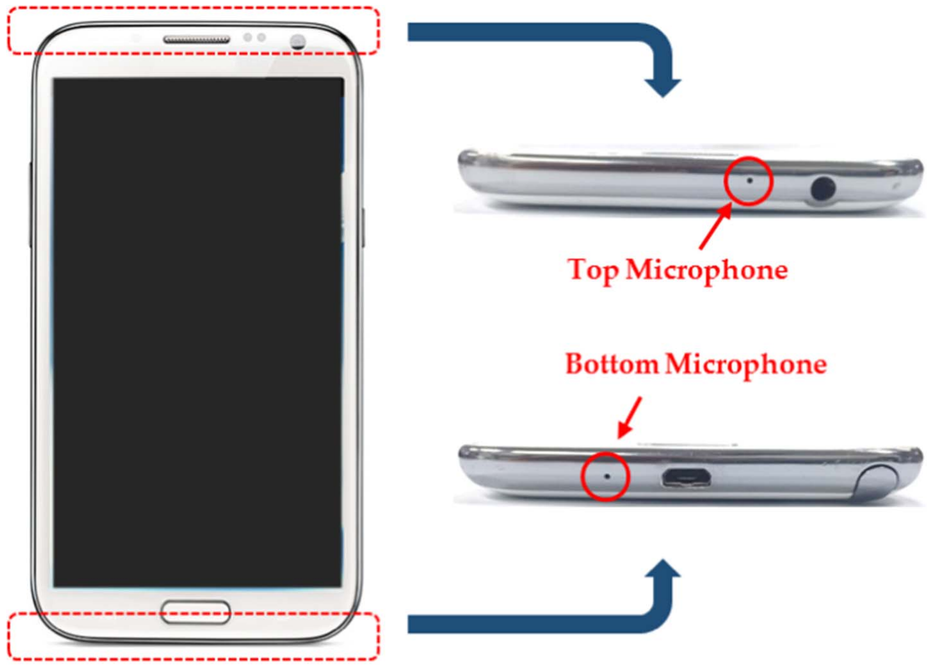

Fig. 1. Two Microphones of a Smartphone.

simulation and measurement. Some possible S-TAP applications are demonstrated in Section 5. Finally, Section 6 summarizes the overall work and draws the conclusions.

2. Related work

2.1. Expanded input modalities

There are several researches about interaction technologies that extend the input modality of mobile devices using embedded sensors.

BackTap [10] is an interaction technique that utilizes tapping motion for input modality. In advance, four distinct tap locations are specified on the back case of a smartphone. Then, by tapping on the back case, a user can interact without occluding the screen by fingers while walking with or holding the smartphone. It employs three common sensors embedded in the smartphone which are a microphone, a gyroscope, and an accelerometer. In this system, the gyroscope and the accelerometer play a main role to detect the tapping events and to classify the tapping locations. In order to further improve the detection performance, a microphone is used as an assistant sensor to capture only the loudness of tapping sounds.

NailSense [11] extracts the position and status of the user's fingernail behind a smartphone by means of a commodity camera. In addition to tracking 2-dimensional (2-D) locations and binary states (touched or not-touched) of fingers, it can also make use of hovering, and bending/extending for the interaction with the smartphone. The color change of the fingernail area is caused due to the tension or pressure when a finger is stretched straight or pressed down. This change enables the smartphone to detect the touched or released state of a finger. Since a finger motion behind a smartphone is in use, users are allowed to control smartphones without occlusion that is one of crucial problems in touch screen applications.

MagGetz [12] employs a magnetometer of a smartphone to provide an additional input modality. It requires user-customizable passive control widgets which can generate the magnetic field without power or wireless connections. This scheme facilitates the tangible interaction. However, magnetic stripes on credit cards and magnetic hard disks may get damaged by widgets and a large offset that must be compensated may occasionally take place at the magnetometer of a smartphone.

Acoustruments [18] is also another passive, power-less, and tangible interaction method which uses a microphone of a smartphone for a continuous ultrasonic sweep from its own speaker. It needs an additional tube attached to a smartphone that is manufactured on multiple methods according to applications. One end of this enclosed tube is connected to the speaker and the other end is directed into the microphone. Like musical instruments, the acoustic properties of ultrasonic sweeps are altered by manipulating structural elements along the tube. Then, the property variation received at a microphone are classified with machine learning algorithms to expand the input modality.

2.2. Audio signal classification

From the perspective of classification of audio signals, AmbientSense [19], RoomSense [20], and audio classification system [21] are related to our work.

AmbientSense is a real-time ambient sound recognition system using a smartphone. In this system, user contexts are recognized by analyzing ambient sounds sampled by a smartphone's microphone and are reported to a user as a real-time feedback. Audio data are sampled at the sampling rate of 16 kHz and framed at a window size of 32 ms with 50% overlap. It uses Mel-frequency cepstral coefficients (MFCC) as features of a classification algorithm that adopts a SVM classifier. The scheme has been evaluated with a data set where daily life ambient sounds are categorized into 23 classes.

RoomSense is an indoor positioning system using a smartphone based on active sound fingerprinting. The acoustic features are extracted from the impulse response measurement and then pattern classification is deployed to recognize room locations within a building floor, similar rooms at different building levels, and different positions within a room. A SVM classifier is incorporated with a Gaussian kernel to discriminate rooms.

Vavrek et al. proposed an audio classification system utilizing a rule-based discrimination and a SVM. It classifies the sound signals into five classes of pure speech, speech with environment sound, speech with music, environment sound, and music. It uses both time domain and frequency domain features entailing a proportion of silent frames, an energy variation in a narrow frequency sub-band, MFCC, and a periodicity of particular frequency sub-band. The rule-based scheme for